EXPERTS’ AND NOVICES’ ABILITIES TO DETECT CHILDREN’S HIGH-STAKES LIES OF OMISSION

Kari L. Nysse-Carris
National Opinion Research Center (NORC) at the University of Chicago

Bette L. Bottoms and Jessica M. Salerno
The University of Illinois at Chicago

We investigated adults’ abilities to detect lies told by 3- to 6-year-old children. Expert forensic interviewers and novices watched videotapes of children who either lied or told the truth about their parent’s transgression, rendered a dichotomous judgment of whether the child lied, rated their confidence in that judgment, and rated the children on various characteristics. Adults detected lies with greater than chance—but not impressive—accuracy, regardless of expertise level. Older children’s lies were more detectable by experts than were younger children’s. Adults were more confident in their judgments about older than younger children. Confidence in lie/truth judgments was not significantly correlated with actual lie detection accuracy.

Keywords: deception detection, legal testimony, lying, child witnesses

The accuracy with which forensic interviewers, attorneys, and jurors judge the veracity of witnesses’ statements can profoundly influence case outcomes. Deceiving and detecting deception are also theoretically interesting topics to social psychologists (e.g., Millar & Millar, 1997; Stiff et al., 1989). In the present study, we compared experts’ (experienced child forensic interviewers’) and novices’ (undergraduates’) abilities to detect younger versus older children’s lies of omission. We improved the ecological validity of previous lie detection research by assessing experts’ and novices’ ability to detect lies told by children who had substantial motivation—protecting their mother from punishment, as they sometimes might in an actual criminal investigation (Bussey 1992).

In some existing lie detection research, children lie about unverifiable emotions or attitudes (e.g., Feldman et al., 1979; Feldman & White, 1980; Morency & Krauss, 1982). Intentionally falsified event reports, however, are of greater
legal relevance than concealed or fabricated emotions (Bottoms et al., 2002). To lie intentionally and successfully about an event, children must alter their story, remember both the actual event and the falsification, and conceal feelings of anxiety, nervousness, or dread associated with deceiving and fear of that deception being detected (DePaulo, 1992). A few researchers have investigated adults’ ability to detect intentionally falsified events. In these studies, adults view children who either lie or tell the truth about events that occur in the laboratory (Edelstein et al., 2006; Goodman et al., 2006; Orcutt et al., 2001). For example, during a laboratory play session, Orcutt and colleagues (2001) had a male “defendant” videotape either children’s (a) exposed arms, stomachs, and toes, or (b) children’s shirt sleeves, belt buckles, and shoes. Later, children were summoned to an actual courthouse to testify in front of mock jurors about this event, in what they believed was an actual trial. Some children were coached by an experimenter to lie by accusing the defendant of videotaping their exposed body parts rather than their clothes. Mock jurors were significantly more likely to convict the defendant when children correctly implicated him than when the children committed perjury or correctly denied that a crime occurred. Jurors also convicted, however, the falsely accused defendant more often than they convicted the defendant who was not guilty. Thus, adults are not necessarily accurate in distinguishing between children’s truthful statements and their lies, a conclusion consistent with results of other studies of children testifying in simulated legal settings (Goodman et al., 2006; Talwar et al., 2006). In fact, adults are no more accurate in detecting child eyewitness’ lies than they are at detecting adult eyewitness’ lies (Edelstein et al., 2006).

The lies of commission in such studies are of greater legal relevance than lies about unverifiable emotions or attitudes, but substantial motivation to lie was still conspicuously missing in these studies. Motivation for lying was limited to complying with an experimenter’s request, and children faced no adverse consequences if their lies were detected. Some motivation is arguably involved in studies using a temptation paradigm, in which children face potential punishment for a transgression. Typically, a child is told not to peek at a toy (and it is inferred that children expect punishment if they peek), is left alone and videotaped to determine whether he or she obeyed, and is then interviewed about whether or not they peeked. Lewis et al. (1989) found that 3-year-old children do peek and lie, and that adults who watched these brief videotaped denials could not distinguish between children who denied actual peeking and children who had not peeked at all. Other studies using a similar paradigm found similar effects. For example, in a study by Leach et al. (2009), undergraduates were unable to detect lies told by 3- to 11-year-old children beyond chance (see also Crossman & Lewis, 2006; Talwar & Lee, 2002).

The Effect of Child Age on Deception Detection

Much research (e.g., Feldman et al., 1979; Feldman & White, 1980; Morency & Krauss, 1982; Westcott et al., 1991), but not all (Newcombe & Bransgrove, 2007; Talwar & Lee, 2002), finds that adults are better able to detect younger than older children’s lies. This could be explained by younger children’s lies being particularly transparent because children only gradually acquire the cognitive and
physiological sophistication necessary for successful concealment of lies about emotions or attitudes through control of nonverbal behaviors (DePaulo & Jordan, 1982; Feldman et al., 1979; Feldman & White, 1980; Saarni, 1984; Saarni, 1990; Shennum & Bugental, 1982; for a review, see Feldman & Philippot, 1991). Talwar et al. (2007) also found that children’s ability to manage verbal behavior associated with successful lying in the peeking paradigm situation increases with age. Thus, nonverbal and verbal behavior regulation generally improves with age. Even so, the extent to which these nonverbal behaviors influence adults’ accuracy in detecting children’s lies is unclear. The studies that found it is easier to detect younger (versus older) children’s lies did not include substantial motivation to lie beyond complying with the experimenter’s request (Feldman et al., 1979; Feldman & White, 1980; Morency & Krauss, 1982; Westcott et al., 1991). The one study, however, that found no effect for child age (Talwar & Lee, 2002) included relatively more motivation for the children to lie—to avoid implied punishment for peeking at a toy. Thus, increasing motivation to lie might ameliorate the effect of age on lie detection ability.

One study compared lie detection ability for younger versus older children’s lies and found that people were more accurate in judging older than younger children’s lies (Newcombe & Bransgrove, 2007). When people were asked to rate the accuracy of two conflicting recounts of a children’s story from same-aged pairs—one accurate and one inaccurate—they performed above chance levels but were more accurate in judging older pairs (i.e., 9-year-olds or adults) compared to younger pairs (4-year-olds). In a second study, when people were asked to judge different-aged pairs, they again performed above chance levels, except when the 4-year-old was accurate and the older person (i.e., 9-year-old or adult) was accurate. These studies, however, have limited applicability to our research question. First, they might be an overestimation of people’s ability to detect lies because they were presented with two conflicting accounts and thus knew that one of the children had to be lying. Even if they did not think either of the targets appeared to be lying, they would guess one of them. In many legal settings, however, people have to make the decision of whether or not to trust one child’s story. Further, this study did not investigate lies of commission rather than omission or high-stakes lies, which are the focus of the current study.

No studies that we know of have addressed lie detection ability in a high-stakes lie situation. In our study, we increased children’s motivation to lie beyond previous temptation paradigm studies in several ways: (a) children were explicitly asked to conceal information by a loved one, (b) children were covering up a significant transgression (breaking a doll they were not supposed to touch), and (c) there was obvious proof of the transgression (i.e., the doll was broken). In so doing, we have included two important aspects of legal settings—high motivation and lies of omission rather than commission (i.e., secret keeping). This situation is relevant to the legal system because adults and children involved in legal investigations sometimes have a variety of motivations for distorting or concealing information, including the desire to avoid punishment, to deny responsibility for wrongdoing, and to protect loved ones (Bottoms et al., 2002; Hoogenraad & McKenzie, 1995; Stouthamer-Loeber, 1991). Also, when children are questioned about crimes, they are often asked to provide incriminating information about loved and trusted adults, such as their parents (Myers, 1998). Although one
concern is that children might falsely accuse innocent adults, another potential and serious problem is that children will lie or conceal information (keep secrets) to protect loved ones or themselves from threatened punishment (Bottoms et al., 2002; Goodwin et al., 1979; Pipe & Goodman, 1991).

Two studies have investigated children’s lies meant to cover a loved one’s transgression. Bottoms et al. (2002) found that children, especially older (5- to 6-year-old) versus younger (3- to 4-year-old) children, will sometimes lie to protect their parent who has broken a toy and asked them to lie about it. Extending that paradigm, Talwar et al. (2004) found that children’s lies are especially likely when they themselves would not be implicated in the transgression. Can people detect children’s lies in these high-stakes situations? Talwar et al. (2006) found that adults were only at chance level in their ability to detect lies told by 4- to 7-year-old children who were coached to lie by their parents during mock trial testimony about events such as attending a wedding—arguably a low-stakes event for which there were no negative consequences for lying. No other studies have addressed the issue of understanding adults’ ability to detect lies told by children who are highly motivated to protect a loved one from punishment.

Motivation to lie might affect older children more than younger children in such circumstances. Compared to 3- and 4-year-olds, 5- and 6-year-olds are generally more cognizant of the meaning of secrecy and are more likely to keep a secret (i.e., lie by omission) to protect their mothers from threatened punishment (Bottoms et al., 2002). Further, children’s moral understanding of lie and truth telling increases as children get older (Bussey, 1992; Talwar et al., 2004; for reviews, see Burton & Strichartz, 1992; Lewis, 1993; Vasek, 1986). Thus, older children, who are more likely to understand the negative consequences of being caught in a lie, might be more motivated liars than younger children. Increased motivation might make their lies more detectable—substantial motivation to lie impairs, rather than facilitates, an adult’s ability to lie successfully (DePaulo et al., 1983; DePaulo et al., 1988) because motivated liars experience anxiety and nervousness when lying, which, in turn, inhibits them from controlling deceptive behaviors in tone of voice and body language (DePaulo, 1992; DePaulo et al., 1983; DePaulo et al., 1988). Following this logic, older children (who might be more motivated to lie) might be more nervous and unable to control deceptive behaviors when compared to younger children. This is a reversal of the direction of the age/detectability effect found in studies investigating the detection of children’s low-stakes lies.

In the current study, adults judged the veracity of 3- to 4-year-old and 5- to 6-year-old children, some of whom had substantial motivation for concealing information: They were explicitly instructed by their own mothers to conceal taboo activities in which the pair had engaged. The reward for lying was to keep the mother from getting in trouble and to receive a prized toy. Thus, our design brings more realism to the investigation of the effect of child age on detection of children’s lies by including high-stakes lies. Based on DePaulo’s (1992) motivation impairment hypothesis, we predicted that adults would detect older children’s lies with greater accuracy than younger children’s lies because older children’s understanding of the consequences of lying and telling the truth would make them more nervous, which would impair their ability to hide their deceptions.
We also predicted that older children would be perceived more positively on global child judgments (i.e., more intelligent, attractive, confident, truthful, likeable, and able to distinguish fact from fantasy). On the one hand, younger children are rated more positively than are older children in cases where children make child sexual abuse allegations because mock jurors perceive young children as truthful, sincere, honest, and cognitively unable to fabricate untrue sexual allegations because they lack sexual knowledge (for review, see Bottoms et al., 2007). On the other hand, younger children are rated as less credible than older children (although more likeable; Bottoms & Goodman, 1994) in other types of eyewitness testimony situations where sexual knowledge is not at issue and cognitive competence is highlighted (Bottoms et al., 2007; Goodman et al., 1984; Leippe et al., 1992). Because our study does not involve child sexual abuse allegations, we hypothesized that older children will be rated more favorably than will younger children—perhaps with the exception of likeability, given Bottoms & Goodman’s (1994) findings.

**The Effect of Expertise on Deception Detection**

In legal settings, children’s truthfulness is judged by people with various levels of expertise, from jurors who are generally novices, to experts trained as forensic interviewers, such as lawyers, police officers, and social workers. Conventional wisdom predicts that professional training, practice, and repeated exposure to legally relevant truths and lies will enhance forensic interviewers’ ability to detect children’s deception. Generally, in fact, experts outperform novices on domain-specific tasks (e.g., chick sexing, Biederman & Shiffrar, 1987; chess playing, de Groot, 1965; see Ericsson & Smith, 1991, for a review). Early lie detection research, however, found no support for expert advantage in lie detection in adults (DePaulo & Pfeifer, 1986; Kraut & Poe, 1980). A recent meta-analysis found no reliable effects of individual differences in peoples’ experience on lie detection ability—lie detection appears to be about the liar and not the perceiver (Bond & DePaulo, 2008). Other work, however, has shown that certain types of expertise offer some advantages (Ekman & O’Sullivan, 1991; Ekman et al., 1999). For example, Ekman et al. (1999) reported that federal law enforcement officers, sheriffs, and forensic psychologists (i.e., professionals with more experience detecting deception) detected lies about adults’ strongly held opinions with somewhat greater accuracy (67% accuracy or higher) than did federal judges and clinical psychologists (i.e., professionals with less experience in detecting deception; 62% accuracy or lower). Ekman and O’Sullivan (1991) found that secret service agents were the only group among other types of experts (i.e., professionals from the Central Intelligence Agency, Federal Bureau of Investigation, National Security Agency, Drug Enforcement Administration, police, and psychiatrists) and laypersons (i.e., college students and working adults) who detected nurses’ lies about their emotional reactions to films beyond chance accuracy. Some experts might excel at lie detection tasks more than others because they either attend to a variety of behavioral cues (Ekman & O’Sullivan, 1991) or they have more training and experience (Ekman et al., 1999).

Research investigating the effect of expertise on adults’ ability to detect children’s lies has been mixed thus far but generally indicates that professionals
who have experience interviewing and judging lies are not more accurate than are novices. For example, in a study by Chahal and Cassidy (1995) examining four 8-year-olds’ deliberately falsified accounts of a film, expertise (as defined by college major: Social Work, Education, Other) did not influence accuracy rates, but parental status did, with parents outperforming nonparents. In a more legally relevant study, social workers and undergraduates viewed interviews ostensibly from sexual abuse investigations, which, in reality, were from mock forensic interviews about an innocuous laboratory event (Goodman et al., 2002). Neither social workers nor undergraduates correctly estimated the children’s actual accuracy (regardless of child age, which ranged from 7 to 10 years), but social workers were significantly less likely to think sexual abuse had occurred than were undergraduates. Finally, Leach et al. (2004) found that undergraduates and custom officers detected children’s lies told within the peeking paradigm at chance level, and police officers actually performed below chance—they tended to label liars as “nonliars” and nonliars as “liars.”

Thus, it appears that expertise is not reliably related to detecting children’s lies. Our ability to draw this conclusion about high-stakes lies, however, is limited because, in those studies, children lacked the substantial motivation to lie. In such circumstances, on the one hand, experts who have worked with children in forensically relevant situations involving these characteristics might be more accurate in their lie detection abilities. On the other hand, such experts might have a general bias to overbelieve children’s reports, which might lead them to misclassify children’s lies as truthful statements. Brigham and Spier (1992) found that, compared to criminal defense lawyers, child advocates (i.e., district attorneys, child protection workers, and police officers) were more likely to characterize children’s reports as accurate and less likely to characterize children’s reports as significantly distorted or completely fabricated. Brigham (1998) found similar pro-child biases in lie detection ability among laypersons. Given that most forensic interviewers are child advocates by choice, and given that they might routinely encounter real cases of child abuse, wherein children are actually abused and give statements that match case evidence (Myers, 1998), experts might hold stronger pro-child biases than laypersons. These pro-child biases might lead them to judge individual children as telling the truth rather than lying about past events, making their lie detection judgments less accurate than those of novices.

Compared to novices, experts might also have greater confidence in their lie/truth judgments because they have extensive experience detecting deception. In fact, experts do express greater confidence in their lie/truth judgments than do novices (e.g., DePaulo & Pfeifer, 1986; Kohnken, 1987; Leach et al., 2004). In the Brigham and Spier (1992) study, 79% of the child protection workers and 76% of the police officers believed that they could usually tell when a child lied. This is problematic because confidence is rarely significantly correlated with accuracy, probably because of overreliance on misleading nonverbal cues (for a meta-analytic review, see DePaulo et al., 1997). Experts’ inflated confidence might lead to even more inaccuracy when combined with a strong pro-child bias.

In the present research, we tested two competing hypotheses regarding the effect of expertise on lie detection. On the one hand, prior research supports a prediction that experts who have extensive and very task-specific training and experience evaluating children’s forensic reports would detect children’s lies with
greater accuracy than would novices who have no experience evaluating children’s forensic reports (e.g., Ekman & O’Sullivan, 1991; Ericsson & Smith, 1991). On the other hand, most studies within the lie detection literature support the alternative prediction that experts would not be more accurate than novices because experts might exhibit stronger pro-child biases in general, which might influence lie detection ability. We predicted that experts would be more confident in their judgments than would novices but that confidence would not be correlated with actual accuracy for either experts or novices. Finally, because experts are likely to encounter real cases of child abuse (Myers, 1998), we predicted experts (versus novices) would rate children more positively on global judgments and be more predisposed to believe children in general.

Method

We investigated, in a high stakes lie situation, whether child age and rater expertise influences adults’ (a) abilities to detect individual children’s deceptions, (b) tendency to label individual children as liars, (c) confidence in lie/truth judgments, (d) global impressions of individual children, and (e) general predisposition to believe all children as a group. The study conformed to a 2 (child age: 3 and 4 years old vs. 5 and 6 years old) × 2 (expertise: novices vs. experts) mixed design, with child age varied within participants and expertise varied between participants. We also investigated whether subgroups of experts who have different levels of experience interviewing children differed in their lie detection abilities. Experts’ years of child interviewing experience ranged from 2 months to 17 years (M = 6 years, SD = 5 years).

Participants

Thirty-six undergraduate psychology students (29 women; age M = 19 years, range = 17–28 years) and 36 experts (29 women; age M = 37 years, range = 24–51 years) participated in the study. Novices and experts were matched for gender and socioeconomic background (predominantly middle class, i.e., modal response = over $80,000 in annual family income). Novices represented diverse ethnic groups: 25% Caucasian, 25% Hispanic, 25% African American, 22% Asian American, and 3% Other. Experts were 83% Caucasian, 11% Hispanic, 3% African American, and 3% Other. Experts were prosecutors (n = 16 or 47%) and nonprosecutors, including clinical psychologists (n = 3 or 9%), police officers (n = 2 or 6%), social workers (n = 13 or 38%), and two others who failed to provide their specific role as psychologist, police officer, or social worker (but who were definitely in one of those three professional roles). In exchange for their participation, novices received class credit; experts received a one-time payment of $10. To motivate participants to attend to the stimulus tape, all had the opportunity to win a $50 gift certificate for being the most accurate.

1 Three additional experts did not complete the study. Specifically, in one experimental session, two experts expressed a desire to receive more information about the circumstances surrounding the videotaped stimuli. Because we were unable to answer their question without compromising the cover story, they withdrew from the study, disrupting the third expert’s participation.
Materials

Stimulus videotape. With parental consent, we used portions of 12 girls’ (six 3- to 4-year-olds and six 5- to 6-year-olds) videotaped interviews from the Bottoms et al. (2002) study as stimuli. For each girl, we included eight question-and-answer pairs concerning a scripted play session involving children and their mothers. Specifically, child-mother pairs and an experimenter entered a laboratory “waiting room” stocked with toys. The experimenter left the room, instructing the pairs in a “secret” condition not to play with the toys, and instructing pairs in a “control” condition that they could play with the toys, while she was gone. As a video camera surreptitiously recorded the pair, all mothers engaged their children in scripted play with the toys, as the mothers had been trained by experimenters to do (unbeknownst to the children). Thus, children in the secret condition were participating in forbidden activities, but those in the control condition were not. All children listened to music, blew bubbles, put on a costume, and posed for a picture. A particularly salient event was when the mother “accidentally” decapitated and hid a Barbie doll. At the end of the play session, mothers in the secret condition reminded their children that “we weren’t supposed to play with these toys,” urged their children to keep the play session “a secret—else I might get in trouble,” and promised their children one of the highly prized toys from the room if they kept the secret. Mothers in the control condition gave no such instructions. Then the experimenter returned and took the child to a new research assistant, who conducted a mock forensic interview about the play session. (For a complete description of the study and the ethical safeguards taken, see Bottoms et al., 2002.)

In the present study, we chose equal numbers of girls from the secret and control conditions. Because control-condition children lacked incentive to lie, except about breaking and hiding the doll, only a few of the younger control-condition children lied during the interview, and the older control-condition children never lied. The control-condition children were included to examine participants’ perceptions of children who truthfully answer questions. That is, it was important to ensure that participants had the opportunity to wrongly judge honest children to be liars. Interspersing video clips of these children among clips of the secret-condition children in the stimulus videotape provided variety among targets rated by participants.

The eight questions used in this study were, in order, (1) “Did you see the scary monster mask?”; (2) “Did you have cookies and juice with your mom in

---

2 One might argue that the experiences of the control and secret children were different given that the secret children were engaged in “forbidden” activities and told to keep them secret, and the control children were not. Actually, all children had high motivation to lie to protect their mothers—at least in response to the two Barbie doll questions—because even the children in the control condition probably assumed that their mothers were wrong to have broken and then hid the doll (or else some children in the control condition would not have lied about it). Further, if anything, the inclusion of control children provides a more conservative test of our hypotheses. Based on DePaulo’s motivation-impairment theory, greater motivation in the group of children being rated (i.e., which might have been the case if all children were in the secret condition) should only result in more differentiation between the age groups, intensifying the effects we report later. In any case, even if the secret and control condition children had somewhat different experiences, those differences were not confounded with our independent variables, so cannot be an alternative explanation for the effects we report.
the waiting room?"; (3) "Did you blow bubbles with your mom when you were waiting?"; (4) "Did you put on a costume?"; (5) "Did your mom turn on the music while you were in the waiting room?"; (6) "Did your mom take a picture of you while you were in the waiting room?"; (7) "Did you see the Barbie doll?"; and (8) "Did anything happen to it?" All of these events happened except having cookies and juice. See Table 1 for the number of lies and truths given in response to these questions.

Separate 2.5-min segments containing the eight question-and-answer pairs from each of the 12 girls’ interviews were professionally edited onto VHS videotape. The order of the girls was determined by a randomized Latin square so that each child’s interview appeared in each rating position once. This produced 12 different tape versions. Within each girl’s interview, after each question-and-answer pair, participants saw a 10-s still frame of the child as she appeared immediately after she answered the question. This allowed participants ample time to make judgments of the child’s truthfulness and their confidence in that judgment (these are the “Individual Item Judgments” described later). Five seconds of blank screen immediately followed this still frame and was accompanied by a tone that alerted participants to look up from their judgment sheets and prepare for the next question-and-answer clip. Thus, within each child’s interview, a 15-s interval separated each question-and-answer pair. A 2-min interval separated each of the 12 children’s interviews, to allow participants to make “global child judgments” about the child they had just seen (described in more detail below). In total, the edited stimulus videotapes were approximately 1 hr long.

Cover story. The cover story provided plausible reasons for the structured nature of the interviews and for the participants to refrain from using any one child’s responses to determine if subsequent children were accurate. (A similar strategy was used by Goodman et al., 1997.) Participants were told that all of the girls were in a previous study, experienced different activities, and that some of their mothers told them to conceal or to fabricate certain events, whereas other mothers told them to tell the truth about certain events. (In reality, as explained previously, all of the child-mother pairs performed the same activities and half of

Table 1

<table>
<thead>
<tr>
<th>Question</th>
<th>Younger (N = 6)</th>
<th></th>
<th>Older (N = 6)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lies</td>
<td>Truths</td>
<td>Lies</td>
<td>Truths</td>
</tr>
<tr>
<td>See scary monster mask?</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Cookies and juice?</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Blow bubbles?</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Put on costume?</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Turn on music?</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>Take a picture of you?</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>See Barbie doll?</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Anything happen to Barbie?</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Totals:</td>
<td>12</td>
<td>36</td>
<td>9</td>
<td>39</td>
</tr>
</tbody>
</table>

Note. See text for full question wording.
the mothers told their children to conceal their activities.) Participants were told that the adult interviewers asked all of the children the same questions and then decided if the children lied or told the truth. Participants were told that the purpose of the present study was to compare whether adults, like themselves, who simply watched the videotaped interviews could identify the children’s lies as well as did the interviewers who were in the room with the children. This purpose was linked to legal debates over the admissibility of videotaped testimony in lieu of live child testimony in a courtroom: Are jurors who watch videotaped testimony as able to discern accuracy as jurors who watch live testimony?

**Rating form.** The rating form provided participants with the age of each child and the transcribed interview question-and-answer pairs. The rating form contained questions assessing participants’ perceptions of the child’s responses to each question-and-answer pair (two Individual Item Judgments) and their overall impressions of each child’s credibility (six Global Child Judgments). The Individual Item Judgments asked participants (a) if the child lied or told the truth and (b) their confidence in that judgment on a 6-point scale ranging from 1 (not at all confident) to 6 (very confident) (similar to prior studies by DePaulo et al., 1982; Feldman et al., 1979; Feldman & White, 1980; Kraut & Poe, 1980). The Global Child Judgments asked participants to rate their overall impressions of the child’s intelligence, attractiveness, confidence, truthfulness, likeability, and ability to distinguish fact from fantasy on six separate 6-point scales ranging from, for example, 1 (very unintelligent) to 6 (very intelligent); 1 (very unattractive) to 6 (very attractive); and so on.

**Children’s General Believability Scale (CGBS).** A general predisposition to believe children was measured with 10 statements answered on a 6-point scale ranging from −3 (strongly disagree) to +3 (strongly agree). Items included (1) “Children usually tell the truth”; (2) “Children frequently tell stories they know are not true”; (3) “Children are basically honest”; (4) “We can’t believe everything that children say because they often confuse fact and fantasy”; (5) “I don’t usually trust what children say unless there’s some kind of evidence to back them up”; (6) “We can’t believe everything kids say because they frequently talk about imaginary animals, monsters, or people as if they are real”; (7) “Children lie a lot”; (8) “I usually believe children”; (9) “I would believe a child more than an adult”; and (10) “Children rarely make up stories to please adults” (adapted from Bottoms, 1993). Coefficient alpha analyses revealed that when all 10 items were included, internal reliability was acceptable (α = .79), but the last two items were dropped for having less than optimal corrected item-total correlations (CITCs). The resulting 8-item scale had acceptable internal reliability (α = .81) and CITCs (range = .43 to .67).

**Demographic questionnaire.** This questionnaire assessed participants’ age, gender, amount of family income, and ethnicity. Experts completed additional questions asking (a) “How long have you been employed as a forensic interviewer of children?”; (b) “About how many years of professional experience with children (other than forensic interviewing) do you have?”; and (c) “Over the course of your professional career, about how many child forensic interviews would you guess you have actually conducted (Circle the grouping that most closely matches the number of interviews you think you personally conducted): 0–49, 50–99, 100–199, 200–299, 300–499, 500–749,
750–999, to 1000+.” The number range of interviews was coded from 1 to 8 for purposes of analyses.

**Procedure**

Experimental sessions were conducted with one, two, or three participants who sat in cubicles (small tables sectioned off with cardboard walls so that participants could not influence each other) arranged in a semi-circle around a 20-in television monitor. The distance from each cubicle to the monitor was approximately 3 ft. Participants were randomly assigned to watch one of the 12 interview orders with two constraints: Three participants in each expertise condition watched the same order, and, within each of the 12 tape orders, experts and novices were matched on participant gender.

After completing informed consent procedures, participants followed along on a written copy of a cover story as the experimenter read it aloud. Then participants were shown one practice interview (one secret condition boy’s set of eight question-and-answer pairs) to orient them to the rating task. Participants had 10 s to make the two Individual Item Judgments for each question-and-answer pair. They repeated this procedure for each of the 12 target children’s interviews for a total of 192 ratings (8 question-and-answer pairs × 2 Individual Item Judgments × 12 children). Following each of the 12 children’s interviews, participants completed the six Global Child Judgments. After all 12 children had been rated and the videotape was over, participants completed the Children’s General Believability and demographic questionnaires. At the end of the session, participants were thanked for their participation, paid or given class credit, and debriefed as per procedures approved by the University of Illinois at Chicago Institutional Review Board.

**Results**

We present results of analyses of proportion accuracy scores first, followed by signal detection theory (SDT) analyses, confidence ratings, global child judgments, and Children’s General Believability Scale scores. SDT analyses are more appropriate for lie detection studies (Swets et al., 1961) because they provide two independent indices of performance: (a) $d'$, deception sensitivity (the ability to identify when a stimulus—in this study, a lie—is present), and (b) $C$, general response bias (the tendency to respond in a particular way when one is uncertain whether a stimulus is present—in this case, the tendency to label statements as lies or as truth when uncertain) (Snodgrass & Corwin, 1988; Swets et al., 1961). A $d'$ of 0 indicates chance; the greater the $d'$ value, the more accurately an observer discriminates lies from truths (Banks, 1970). Unbiased responses produce $C$ values of 0, liberal biases (i.e., being prone to suspect lying) have negative $C$ values, and

---

3 Parametric and nonparametric indices of sensitivity and bias can be calculated using ordinates defined by the false alarm and hit rates. Parametric indices have strict homogeneity and normality assumptions but are more sensitive than nonparametric indices (Grier, 1971). Snodgrass and Corwin (1988) demonstrated that nonparametric indices do not yield independent measures of sensitivity and bias (see also Banks, 1970).
conservative biases (i.e., being prone to suspect truthfulness) produce positive $C$ values (Snodgrass & Corwin, 1988). We also present proportion accuracy scores, even though they confound sensitivity with bias (Banks, 1970), because they allow for comparisons to other studies whose authors reported only proportion accuracy scores and not SDT analyses.

**Main Analyses**

Unless otherwise noted, all dependent measures were analyzed with 2 (child age: 3- to 4-year-olds, 5- to 6-year-olds) $\times$ 2 (expertise: novice, expert) mixed-groups ANOVAs with child age varied within participants and expertise varied between participants.4

**Proportion accuracy scores.** Proportion accuracy scores were calculated by dividing the total number of lies and truths correctly identified by the total number of dichotomous lie/truth judgments made. For example, if a participant correctly identified 9 lies and 18 truths while making 48 lie/truth judgments for a particular child age group, his or her proportion accuracy score was 56%.5

As expected, adults were more accurate when judging answers given by older than younger children, $F(1, 70) = 6.74, p < .05$, partial $\eta^2 = .09$ (see Table 2). Neither the main effect of expertise, $F(1, 70) = .08$, partial $\eta^2 = .00$, nor the interaction of child age and expertise was statistically significant, $F(1, 70) = .80$, partial $\eta^2 = .01$.

---

4 Compared to novices, experts were significantly older, $t(68) = 14.16, p < .001$, and more likely to be parents, $\chi^2(1) = 10.67, p < .01$; Ms = 8% and 42%, respectively. In prior research, participant age has not influenced the ability to detect deception (e.g., Ekman et al., 1999), but parental status has (e.g., Chahal & Cassidy, 1995). Exploratory analyses comparing the 18 parents and 54 nonparents found that parents expressed the strongest predisposition to believe children, $F(1, 69) = 12.45, p < .01$, but their $d'$ scores ($M = .48$) were not significantly different from nonparents' $d'$ scores ($M = .46$), $F(1, 69) = .02, p = .88$. In any case, whenever a significant effect of expertise emerged in our main analyses, we included parental status as a covariate in a subsequent analysis of covariance (ANCOVA). These additional ANCOVAs revealed only one finding that differed from the results we report. Specifically, when parental status was included as a covariate in the analysis of attractiveness ratings, the expertise effect became nonsignificant, $F(1, 69) = 2.96, p = .10, \eta^2 = .04$. Parental status, however, was not a significant covariate in that analysis, $F(1, 69) = .12, p = .73$.

5 A post hoc analysis was conducted to determine if (a) rater accuracy improved (i.e., a practice effect) or declined (i.e., a fatigue effect) over the course of rating the interview tape, and (b) if there was an effect, if it was similar across all conditions. We compared proportion accuracy scores for the first and second halves of the rating tape (i.e., the first six children vs. the last six children). Proportion accuracy scores were entered into a 2 (tape half: first half vs. second half) $\times$ 2 (child age) $\times$ (expertise) mixed-groups ANOVA, with tape half and child age varied within subjects and expertise varied between subjects. Proportion accuracy scores were used because $d'$ scores for both tape halves could not be calculated for every participant. That is, $d'$ scores could be calculated only for children who lied, and it was possible for all of the liars in a given age group to appear in the same tape half (i.e., the other half of the tape lacked a representative liar from the age group). Adults' proportion accuracy scores significantly improved from the first half of the tape ($M = .61$) to the second half of the tape ($M = .66$), $F(1, 70) = 11.31, p < .01$; however, no statistically significant interactions emerged, $Fs(1, 70) \leq 2.58, ps > .11$. Because practice effects were evenly distributed across child age—each child appeared once in each rating position of the randomized Latin square—tape half was not included as a variable in any subsequent analyses.
analyses. Participants’ $d'$ scores for both younger and older children exceeded chance (defined as $d' = 0$), $t(71) \geq 6.71, ps < .001$. Lies told by older children were more detectable ($M = .38, SD = .48$) than lies told by younger children ($M = .33, SD = .56$), $F(1, 70) = 6.48, p < .05$, partial $\eta^2 = .09$ (see Table 2), but this effect was qualified by a significant child age by expertise interaction, $F(1, 70) = 4.12, p < .05$, partial $\eta^2 = .06$. Simple effects analyses revealed that experts were more sensitive to older than younger children’s lies, $F(1, 70) = 10.46, p < .01$, partial $\eta^2 = .22$, but novices were not affected by child age, $F(1, 70) = .14, ns$, partial $\eta^2 = .004$. There was no significant main effect of expertise, $F(1, 70) = .05, ns$, partial $\eta^2 = .00$.

Confidence ratings. Mean confidence-in-lie/truth judgments were calculated by averaging ratings made within each child age group. Thus, each participant contributed 48 ratings to the mean for each age group (i.e., 6 children in each age group × 8 ratings made for each child). Adults were more confident about their judgments made for older ($M = 4.34, SD = .50$) than younger children ($M = 4.16, SD = .49$), $F(1, 70) = 31.71, p < .001$, partial $\eta^2 = .31$ (see Table 2). Contrary to hypotheses, experts were not significantly more confident in their judgments than were novices, $F(1, 70) = 1.28, ns$, partial $\eta^2 = .02$. The child age by expertise interaction was not statistically significant, $F(1, 70) = .02, ns$, partial $\eta^2 = .00$.

Correlations among confidence ratings and accuracy measures. We calculated Pearson correlations between participants’ overall mean confidence in their lie/truth judgments and (a) overall proportion accuracy scores based on judgments for all 12 target children, (b) proportion accuracy scores for younger and older children separately based on judgments for all six children in each age group, (c) overall $d'$ score based on judgments made for all 12 target children, and (d) the $d'$ scores for younger and older children separately based on all six children in each age group. (See Table 3 for all correlations.)
As expected, novices’ and experts’ confidence in their lie/truth judgments was a poor indicator of actual accuracy (i.e., proportion accuracy and $d'$ scores) in detecting lies told by younger children, all $r$s $\leq .10$, $ns$; older children, all $r$s $= .16$, $ns$; and younger and older children combined, all $r$s $= .18$, $ns$.

**Bias scores.** Overall, negative $C$ bias scores for both age groups significantly differed from zero, indicating bias, $t(71) = 3.69$, $p < .001$: Participants had a tendency to label children “liars” when uncertain about their truthfulness. As expected, there were no differences in this tendency for older ($M = .19$, $SD = .44$) versus younger children ($M = -.25$, $SD = .39$), $F(1, 70) = 1.29$, $ns$, partial $\eta^2 = .02$ (see Table 2). Neither the main effect of expertise, $F(1, 70) = .10$, $ns$, partial $\eta^2 = .03$, nor the child age by expertise interaction were statistically significant, $F(1, 70) = 2.14$, $ns$, partial $\eta^2 = .03$.

**Global child judgments.** Six separate $2$ (child age: 3 to 4 years old, 5 to 6 years old) $\times 2$ (expertise: forensic interviewers, novices) ANOVAs revealed that, as expected, adults judged older children to be more truthful, intelligent, attractive, confident, likeable, and able to distinguish fact from fantasy than younger children, $F$s$(1, 70) = 8.72$, $p < .01$, partial $\eta^2$s $\geq .11$ (see Table 4). Expertise, however, did not influence these Global Child Judgments, $F$s$(1, 70) = 1.73$, $p \geq .19$, $\eta^2$s $\leq .03$, with one exception: Experts rated children as more attractive than did novices, $F(1, 70) = 4.08$, $p < .05$, partial $\eta^2 = .06$. No child age by expertise interactions were significant, $F$s $(1, 70) = 1.81$, $ps \geq .18$.

**Children’s General Believability Scale scores.** A one-way ANOVA revealed that, compared to novices ($M = 3.72$, $SD = .62$), experts ($M = 4.47$, $SD = .62$) expressed a stronger overall predisposition to believe children in general, $F(1, 69) = 26.30$, $p < .001$, partial $\eta^2 = .26$. Pearson correlations were conducted to examine the separate relations between Children’s General Believability Scale scores and proportion accuracy, $d'$, $C$, and Global Child truthfulness ratings. No statistically significant correlations emerged, all $r$s $\leq .20$, $ps \geq .10$. Thus, participants’ overall predispositions were independent of their specific ratings of the children in this study.

**Exploratory Analyses Explaining the Null Effects of Expertise**

We did not find an expertise effect on detection accuracy (although experts were more sensitive to older than younger children’s lies, whereas novices were
Because Ekman and colleagues (1991, 1999) have argued that nonsignificant expertise effects might be due to a failure to examine individual differences among the types of experts composing an expert group, we conducted exploratory post hoc analyses to determine whether lie detection ability differed among our expert subgroups, specifically, prosecutors versus nonprosecutors. One could speculate that clinical psychologists, social workers, and police officers might have more daily experience in face-to-face interactions and formal interviews with children than attorneys, which might lead them to be more accurate.

So we compared prosecutors, who represented the single largest profession ($n = 16$ or 47%), to nonprosecutors ($n = 20$, 53%). Separate 2 (child age: 3 to 4 years old, 5 to 6 years old) $\times$ 2 (expert type: prosecutor, nonprosecutors) mixed-groups ANOVAs were performed on proportion accuracy, $d'$, and $C$ scores. Prosecutors identified children’s statements with significantly greater proportion accuracy ($M = .66$, $SD = .10$) than did nonprosecutors ($M = .60$, $SD = .07$), $F(1, 34) = 4.89$, $p < .05$, partial $\eta^2 = .13$, and prosecutors were more sensitive to children’s lies ($d'$) ($M = .75$, $SD = .69$) than were nonprosecutors ($M = .29$, $SD = .55$), $F(1, 34) = 7.04$, $p < .05$, partial $\eta^2 = .17$. Prosecutors and nonprosecutors did not significantly differ on $C$ bias scores ($Ms = -.18$ and $-.23$, $SDs = .54$ and .41, respectively), $F(1, 34) = .11$, $ns$, partial $\eta^2 = .001$.

To determine what accounted for expertise differences, we conducted mediation analyses using regression models (Baron & Kenny, 1986) to determine whether three potential mediators explained the expert type effect: (a) years employed as a forensic interviewer of children, (b) years of other (i.e., nonforensic) professional experience with children, and (c) the estimated total number of child forensic interviews conducted. The two expert groups did not differ on years of forensic interviewing nor nonforensic professional experience with children.

### Table 4

**Mean Global Child Judgments as a Function of Child Age and Expertise**

<table>
<thead>
<tr>
<th>Child age</th>
<th>Novices</th>
<th>Experts</th>
<th>$M$</th>
</tr>
</thead>
<tbody>
<tr>
<td>3- and 4-year-olds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truthfulness</td>
<td>3.69 (.54)</td>
<td>3.75 (.51)</td>
<td>3.72</td>
</tr>
<tr>
<td>Intelligence</td>
<td>4.07 (.51)</td>
<td>4.07 (.45)</td>
<td>4.07</td>
</tr>
<tr>
<td>Attractiveness</td>
<td>4.02 (.68)</td>
<td>4.35 (.50)</td>
<td>4.19</td>
</tr>
<tr>
<td>Confidence</td>
<td>3.64 (.53)</td>
<td>3.58 (.44)</td>
<td>3.61</td>
</tr>
<tr>
<td>Likeability</td>
<td>4.26 (.50)</td>
<td>4.44 (.49)</td>
<td>4.35</td>
</tr>
<tr>
<td>Ability to distinguish fact/fantasy</td>
<td>3.23 (.73)</td>
<td>3.38 (.71)</td>
<td>3.31</td>
</tr>
<tr>
<td>5- and 6-year-olds</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Truthfulness</td>
<td>3.94 (.66)</td>
<td>3.91 (.53)</td>
<td>3.93</td>
</tr>
<tr>
<td>Intelligence</td>
<td>4.58 (.48)</td>
<td>4.62 (.46)</td>
<td>4.60</td>
</tr>
<tr>
<td>Attractiveness</td>
<td>4.31 (.57)</td>
<td>4.47 (.51)</td>
<td>4.39</td>
</tr>
<tr>
<td>Confidence</td>
<td>4.20 (.53)</td>
<td>4.13 (.49)</td>
<td>4.17</td>
</tr>
<tr>
<td>Likeability</td>
<td>4.55 (.53)</td>
<td>4.70 (.41)</td>
<td>4.63</td>
</tr>
<tr>
<td>Ability to distinguish fact/fantasy</td>
<td>4.07 (.63)</td>
<td>4.30 (.74)</td>
<td>4.19</td>
</tr>
</tbody>
</table>

*Note.* Standard deviations shown in parentheses. Ratings were made on 6-point scales where higher values indicate more of the attribute.
Prosecutors did, however, conduct significantly fewer interviews than did nonprosecutors, $\beta = -0.40$, $t(34) = -2.56$, $p < .05$. When expert type and number of interviews conducted were entered simultaneously as predictors, the number of interviews conducted did not significantly mediate the expert type effect. Although the beta for expert type was reduced slightly from 0.41 to 0.37, it was still significant, $t(33) = 2.14$, $p < .05$, and the number of interviews was not a significant predictor of $d'$, $\beta = -0.12$, $t(33) = -0.68$, ns.

Thus, as suggested by Ekman and colleagues (1991, 1999), different kinds of experts may excel more than others at lie detection tasks. In our case, prosecutors were at an advantage compared to nonprosecutors, but this difference was not explained by years of child forensic interviewing experience, years of nonforensic professional experience with children, or the number of child forensic interviews a professional had conducted.

**Discussion**

Although most studies find that detection is at or near chance levels (e.g., Crossman & Lewis, 2006; Leach et al., 2009; Lewis et al., 1989; Talwar & Lee, 2002), in our study, adults detected children’s high-stakes lies of omission with significantly greater-than-chance accuracy. This is a noteworthy finding considering the literature’s “virtually axiomatic” conclusion that, on average, people are not able to detect lies beyond chance (Bond & DePaulo, 2008, pp. 477, 485). We believe this is because the children in our study had greater motivation to lie than in previous studies, suggesting that it is easier to detect children’s high-stakes lies than lies with less serious consequences. This interpretation is consistent with DePaulo’s motivation impairment hypothesis, which predicts greater motivation to lie would make children more nervous and, in turn, impair their ability to hide their deceptions. Thus, prior work might have underestimated the detectability of legally relevant lies, which are usually by nature high-stakes lies.

In contrast to the overall truth bias reported in previous literature (Bond & DePaulo, 2006), we found that when they were uncertain, participants were biased to label children as liars—regardless of child age or participant expertise. This might be another result of studying high-stakes lies. That is, consistent with our points above, perhaps so many children on the tapes leaked their nervousness as to make participants somewhat suspicious of all the children, biasing the observers to suspect more lying than participants typically suspect in studies of low-stakes lying. It is also possible that this overall suspicion resulted from the fact that children were asked closed-ended questions, which tend to result in lower ratings of credibility than do open-ended questions that prompt longer answer narratives (e.g., Tubb et al., 1999). Finally, our cover story highlighted the possibility that all children could lie during the interview, perhaps priming suspicions of lying more than truth telling.

---

6 Westcott et al. (1991) did find that, overall, adults’ proportion accuracy was 59%. Although this proportion was technically significantly different than chance (50%), they did not conduct signal detection analyses, and, as the authors point out, in practice, this is not a meaningfully significant improvement beyond chance.
Confidence in lie/truth judgments was a poor predictor of adults’ accuracy. Why? Social psychological research indicates that people are overly confident in the correctness of their judgments, regardless of their accuracy (the “overconfidence effect,” e.g., Dunning et al., 1990; Kahneman & Tversky, 1996). Further, this overconfidence effect is particularly characteristic of difficult tasks, such as lie detection (e.g., Lichtenstein & Fischhoff, 1977; DePaulo et al., 1997). As outlined by DePaulo and her colleagues (1997), adults’ theories of deceptive behavior might lead them to feel very confident when they observe behaviors they associate with lying. Unfortunately, those behaviors might not always mark deception.

Child Age Effects

Overall, people were better able to detect lies told by 5- and 6-year-olds than lies told by 3- and 4-year-olds, although the size of this main effect was small and the effect held only for experts (with a reasonable effect size). Why were older children easier to detect than younger children in our study, when all previous work found either the opposite (e.g., Feldman et al., 1979; Feldman & White, 1980; Westcott et al., 1991) or no age effect (Talwar & Lee, 2002)? Previous research and developmental theory suggest that adults should detect younger children’s lies with greater accuracy than older children’s lies because children gradually develop the ability to regulate nonverbal behaviors (e.g., Feldman et al., 1979; Westcott et al., 1991). This effect, however, might depend on how motivated children are to lie. Perhaps older children, who were more cognizant of the meaning of secrecy—as indicated by their increased tendency to keep secrets and increased understanding of secrecy (Bottoms et al., 2002)—experienced greater anxiety and apprehension when they were pointedly questioned about the taboo play session. According to DePaulo’s motivation impairment hypothesis (DePaulo et al., 1983, 1988), motivated liars’ nonverbal behaviors are particularly revealing of their deceptions. Thus, adults might have been appropriately influenced by the older children’s nonverbal behaviors, which increased their ability to detect the older children’s lies. Younger children—who were less sophisticated in terms of their ability to keep secrets (Bottoms et al., 2002)—might not have experienced as much apprehension and worry as did older children, and, therefore, their nonverbal behaviors might not have leaked their deception as clearly as the older children’s did. Thus, the benefit older children have over younger children when telling lies of lesser consequence might be reversed in a high-stakes lie situation.

Expertise Effects

Consistent with prior research (e.g., DePaulo & Pfeifer, 1986; Goodman et al., 2002; Kraut & Poe, 1980), there were no significant main effects of expertise on adults’ ability to detect children’s deceptions, as measured by proportion accuracy and $d'$ scores, on adults’ overall confidence in the correctness of their lie/truth judgments, or on their tendency to label children liars (as measured by $C$ scores). Experts (but not novices), however, detected older children’s lies more accurately than younger children’s lies. Also, experts were more likely to be predisposed to believe children than were novices (as measured by the CGBS). Expertise might
not influence the ability to detect lies from a child’s demeanor, but it might influence other skills related to lie detection, such as catching inconsistencies, asking probative questions, and setting dragnets (Kraut, 1980). Of course, our experimental design precluded the experts’ access to any such cues or any opportunity to ask questions themselves, something they in fact complained about. The experts we used also did not receive consistent outcome feedback as to the correctness of their judgments in their work.

Even so, expertise did have some effect. Supporting Ekman et al.’s (1999) contention that some experts might be more accurate than others due to their more extensive experience, and consistent with prior research (Ekman & O’Sullivan, 1991; Ekman et al., 1999), prosecutors detected children’s lies with greater accuracy than did nonprosecutors, an effect that was not mediated by the most obvious explanation: prior experience in interviewing children. Of course, these analyses were post hoc in nature, and we believe that a controlled study aimed specifically at this issue is necessary before drawing definitive conclusions.

Contrary to hypotheses and prior research (e.g., DePaulo & Pfeifer, 1986), experts were not more confident in the correctness of their lie/truth judgments than were novices. The demands of the experimental task might explain these discrepant results. Specifically, experts recognized that the lie detection task did not mirror real-life forensic interviews, whereas novices were probably less aware of this. For example, when asked how she thought her lie detection performance would compare to a novice, one expert noted that because “this is not a clinical interview . . . (I would perform) perhaps the same as the ‘next guy.’” One of the novice participants noted, “(Experts are) trained to tell. I was not too confident I chose the right answer.”

Expertise did not influence bias scores (measured by $C$), with experts and novices equally likely to label children “liars.” As we mention above, our cover story highlighted the possibility that all children could lie during the interview, and this might have overpowered any predispositions tied to expertise that might have led participants to judge children to be truth tellers or liars (e.g., Ross & Nisbett, 1991).

**Caveats, Implications, and Future Directions**

Caution must be exercised when attempting to generalize these results to actual forensic interviewer lie detection situations—the lie detection task lacked some mundane realism. For example, experts often conduct face-to-face interviews—they typically do not observe videotaped interviews to make veracity judgments, although they sometimes do. Thus, the experimental task employed here might have failed to fully test their expertise. A more realistic test would entail face-to-face interviews in which experts and novices freely question children and children give longer narratives. Such a design, however, would involve costly trade-offs of experimental control, which we retained in our design.

Further, our results might not generalize to all interviewing situations, such as instances when children fabricate events (as opposed to denying events, as in our study). Thus, replication is certainly important. Future studies could also address a number of other interesting factors we were unable to study. For example, boys appear to lie differently than girls do, with boys leaking more deception cues than
girls (e.g., Feldman et al., 1979; Saarni, 1984), but do boys and girls who have compelling motivation to lie differ in their ability to lie and avoid detection by adults? Would such an effect be moderated by age or by observer gender (which, it should be noted, was imbalanced in our study, with more women than men participants)?

Our research adds to the basic social psychological and developmental literature on adults’ ability to detect deception and children’s ability to deceive others by highlighting the importance of considering substantial motivation to lie (lies told to protect a parent). Because our investigation of adults’ ability to detect children’s lies focused on high-stakes lies of omission told to protect a parent, we uncovered a unique situation under which people are able to detect children’s lies beyond chance. This is of considerable importance to basic research regarding lie detection, from which has been drawn the opposite conclusion—that people are not able to detect lies beyond chance (Bond & DePaulo, 2008).

Our study is among the most ecologically valid of such studies because the children’s lies were driven by a realistic and compelling motive (i.e., to protect themselves and their mothers from punishment). Our lying situation also represents well the legal setting of initial forensic investigations because the children were telling lies of omission versus commission. That is, an important task often faced by child maltreatment investigators (e.g., police officers, social workers, forensic interviewers, and attorneys) is to determine whether a child is denying actual crime, such as physical or sexual abuse, that he or she is reluctant to disclose, versus offering true denials of abuse that never happened. Results indicate that although this is indeed a difficult task, investigators may do this with better-than-chance accuracy even in a highly controlled laboratory situation, and may be especially accurate when children are older than 4 years. Future studies that allow investigators the tools they use each day in their jobs (e.g., access to contextual information, ability to ask follow-up questions freely) might reveal even higher accuracy.

It is also worth noting that an applied implication of the expertise by age interaction is that experts should not assume that younger and older children will be affected similarly by high-stakes lying situations. Younger children might not be able to understand the high-stakes consequences of their lies, and might therefore not leak emotion as much as older children if they lie. It is conventional wisdom that younger children are less effective liars, which could potentially make experts less diligent at detecting younger versus older children’s lies.

In summary, our study represents a valid first step in attempting to understand adults’ ability to detect children’s deceptions, and our findings are helpful in understanding how accurate experts might be in assessing the accuracy of children interviewed in forensic contexts.

References


Received April 11, 2010
Revision received October 12, 2010
Accepted October 13, 2010 ■